

Detecting Cognitive Appraisals from Facial Expressions for Interest Recognition

Mohammad Soleymani

Swiss Center for Affective Sciences

University of Geneva

Switzerland mohammad.soleymani@unige.ch

Abstract—Interest makes one hold her attention on the object of interest. Automatic recognition of interest has numerous applications in human-computer interaction. In this paper, we study the facial expressions associated with interest and its underlying and closely related components, namely, coping potential, novelty and complexity. We develop a method for automatic recognition of visual interest in response to images and micro-videos. To this end, we conducted an experiment in which participants watched images and micro-videos while their frontal videos were recorded. After each item they self-reported their level of interest, coping potential and perceived novelty and complexity. We used OpenFace to track facial action units (AU) and studies the presence of AUs with interest and its related components. We tracked the facial landmarks and extracted features from each response. We trained random forest regression models to detect the level of interest, curiosity, and appraisals. We obtained promising results on coping potential, novelty and complexity detection. With this work, we demonstrate the feasibility of detecting cognitive appraisals from facial expressions which will open the door for appraisal-driven emotion recognition methods.

I. INTRODUCTION

Interest drives our focus of attention. It is related to novelty, complexity, familiarity and unrelated to pleasantness [1]. Research in psychology suggests that interest satisfies the conditions for being an emotion. It has an appraisal structure and bodily expressions [2] and is mediated by personality [3]. Appraisals are a set of cognitive evaluations in response to an event or object that are important in the construction of emotions [4]. Silvia identified the appraisals of coping potential and novelty-complexity to be the main driving appraisals for interest. Automatic recognition of the expressions associated with interest has numerous applications, from education and learning [5] to content recommendation and retrieval [6]. The research in psychology identified a number facial movements that are active when a person is interested, including eyelid widening and parting lips [7]. Both these facial movements are also associated with the appraisal of novelty and coping potential (AU5 and AU 25) [8]. Mortillaro *et al.*[9] proposed that emotion recognition should be done through recognizing cognitive appraisals. If we recognize the appraisals as constructing factors of emotions, we can move beyond the current methods which are mainly based on the automatic recognition of prototypical emotions.

This work was supported by Swiss National Science Foundation Ambizione grant and in part by Biotech campus foundation

In this work, we aim at detecting interest and appraisals associated with interest, namely, coping potential and novelty-complexity [2]. We first selected a diverse set of images and micro-videos covering a wide range of topics and emotional content. We then recorded a dataset of spontaneous responses from 50 participants watching and rating the content. Participants reported the level of interest, curiosity, coping potential (comprehension level), perceived novelty and complexity after each stimulus. Facial expressions were analyzed by tracking landmarks and facial action units (FAU). We found significant correlations between the present facial action units and appraisals and interest. After registering faces to a normal face, we extracted features from each frame and pooled them for each trial to form a feature vector. We trained an ensemble regression model, random forest, for detecting appraisals, curiosity and interest from facial expressions. The results demonstrate the ability of facial expressions in capturing patterns associated with appraisals and interest. We then combined multiple participants responses for automatic detection of general interestingness of the content. In summary the major contributions of this paper are:

- For the first time, we report on detecting facial expressions associated with cognitive appraisals.
- We provide an analysis of the expressions associated with interest and appraisals.

In the remaining of this paper, we will familiarize the reader with the existing work, present our material and method, draw conclusions and provide perspectives.

II. BACKGROUND

Silvia [3], [1] has studied the appraisal mechanism of interest. He found novelty-complexity and coping potential to be the most important appraisals in the process of feeling interest. Coping potential is the ability to cope with an event, for example, in case of images, Silvia used comprehensibility for assessing coping potential [1]. He has also identified that people with a higher level of familiarity with the subject have a higher level of interest in more complex forms of the stimuli. He later found that people can be categorized into different groups regarding how they feel interest towards an object or situation [3]. The first group, with a higher curiosity personality trait, are more likely to be interested by novelty and more complex stimuli. For the second group, however, coping potential and comprehensibility was more important.

It is also important to note that interest is not always co-occurring with pleasant emotions and there are unpleasant experiences that might elicit interest [1].

Users' interest in content can help recommender systems, content producers and advertisers to better focus their efforts towards higher user satisfaction. Reeve [7] studied the facial movements as well as physiological responses during the experience of interest. He showed a set of interesting and non-interesting videos to the participants of his experiments. He has identified a set of facial expressions, eye gaze behavior and head movements, such as head stillness and parting lips, that are associated with interest. Kapoor and his colleagues [10], [5] used game state, body posture, facial expressions and head pose to detect interest in children playing an educational game. Body posture was sensed by a grid of pressure sensors installed on the chair where the child was sitting. They could accurately detect interesting situations during the game play. Body posture was the most informative modality for interest detection. Gatica-Perez et al. [11] proposed a system to recognize the level of interest in a group meeting from audiovisual data. A dataset of audiovisual recordings from scripted or posed meetings was annotated for the moments of interest, e.g., the moments that people were attentive and took notes. Audiovisual features were extracted from the participants' faces and voices. The audio channel was the most informative modality in their setting and dataset. Kurdyukova et al. [12] set up a display that could detect the interest of the passersby by detecting their faces, facial expressions and head pose. The most comprehensive study on recognition of interest was done by Schuller et al. [13]. They recorded an audiovisual interest corpus (AVIC). In their experiment, the experimenter and the participant were sitting on opposite sides of a table. The experimenter played the role of a marketer presenting a product to the participant. The participant was encouraged to engage in a conversation and ask questions. Audiovisual data were recorded and the segmented speaker and subspeaker turns were annotated by the degree of interest on a five points scale. The five degrees of interest were from disinterest to curiosity. Speech and non-linguistic vocalizations were transcribed and labeled by human transcribers. Across different modalities, acoustic features were shown to perform the best.

III. MATERIAL

A. Stimuli content

In a preliminary study [14] 1005 pictures were selected from Flickr¹ covering various topics. Pictures received 20 labels on interestingness, comprehensibility (coping potential), pleasantness, aesthetics arousal, complexity and novelty on Amazon Mechanical Turk². 80 images were selected to cover the whole spectrum in terms of average interestingness, pleasantness and coping potential.

132 micro-videos in GIF format from Video2GIF dataset [15] were randomly selected and annotated on similar scales

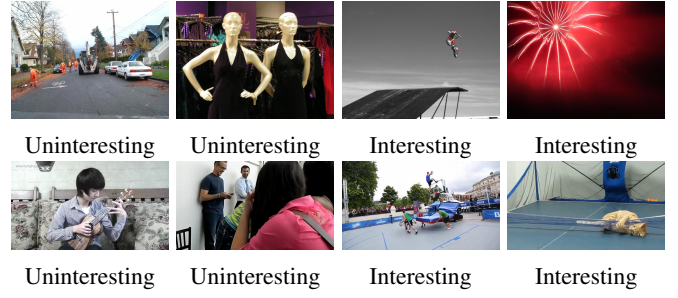


Fig. 1: Examples of stimuli content; the first row shows examples of images and the second row snapshots from micro-videos.

on Amazon Mechanical Turk. The Video2GIF GIFs are extracted from YouTube videos. As a result, we could obtain the higher quality equivalent videos and re-encode them to our desired format. 40 micro-videos were selected to cover the whole spectrum in terms of average interestingness, pleasantness and coping potential. Since in the experiment we were interested in using the GIF like characteristics of the clips, we re-encoded the videos with 1.5x speed and repeated the sequence twice to demonstrate the possible loopiness. Microvideos were in average 11 seconds long. Examples of the stimuli are given in Fig. 1.

B. Recordings

The experiment has received ethical approval from the ethics commission of the faculty of psychology and educational sciences, University of Geneva. 52 healthy participants with normal or corrected to normal vision were recruited through campus wide posters and Facebook. From these 52 participants, 19 male 33 female. Participants were in average 25.7 (*standard deviation* = 5.3) years old. Participants were informed about their rights and signed an informed consent form before starting the experiment. They received a monetary gratitude for their participation. They were first familiarized with the protocol and ratings, in a dummy run. Experiments were conducted in an acoustically isolated experimental booth. Video was recorded using an Allied Vision³ Stingray camera at 60.03 frames/second and with 780x580 resolution. Two Litepanels⁴ daylight spot LED projectors were used for lighting. Video was recorded by Norpix Streampix software⁵. Protocol was ran by Tobii Studio software and it was synchronized by a sound trigger that marked the frames before each stimulus. To simplify the interface we only provided the participants with a keyboard with numerical buttons that they could use to give ratings (1-7). A picture of the experimental setup is shown in Fig. 3. We have also recorded eye gaze and galvanic skin responses. In this paper, we only report the analysis on facial expressions. Examples of facial expressions are given in Fig. 2.

¹<http://www.flickr.com>

²<http://www.mturk.com>

³<https://www.alliedvision.com/>

⁴<http://www.litepanels.com>

⁵<https://www.norpix.com>

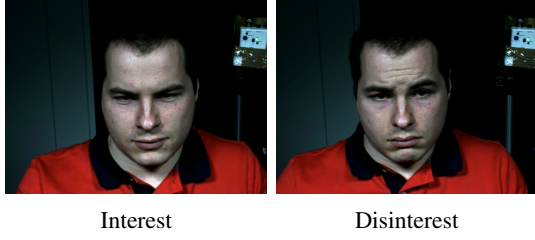


Fig. 2: Examples of expression in two cases of low and high interest.



Fig. 3: The recording setup including an eye gaze tracker, camera capturing front face video and galvanic skin response.

Participants looked at images for five seconds and rated them on their interestingness, invoked curiosity, perceived coping potential, novelty and complexity on a seven point semantic differential scale. Interestingness was assessed by rating from uninteresting to interesting. Curiosity was assessed by asking how much do they like to watch or look at similar content. Coping potential was assessed by averaging the ratings given to the item on two closely related scales; easy to understand to hard to understand and incomprehensible to comprehensible. Novelty was assessed by rating the items from not novel to novel. Complexity was assessed on simple to complex. The correlation coefficient between different ratings of 120 items (40 micro-videos and 80 images) and their inter-rater agreements are given in Table I. As expected interest and curiosity have a very high correlation.

IV. FACIAL EXPRESSIONS OF INTEREST AND APPRAISALS

The data from 2 participants had to be discarded due to the technical failure in recording the videos and synchronization. Head pose, head scale and eye gaze coordinates were extracted in addition to the facial action units. The intensity of the following action units were detected at frame level by OpenFace [16], [17]: AU1, AU02, AU4, AU5, AU6, AU7,

TABLE I: Krippendorff’s alpha inter-rater agreement on ordinal scale and Spearman correlation coefficient between the ratings.

scale	interest	Coping	Curiosity	Novelty	Complexity
Interest	-	0.067	0.77	0.31	0.26
Coping	-	-	0.00	0.39	0.67
Curiosity	-	-	-	0.27	0.19
Novelty	-	-	-	-	0.44
Complexity	-	-	-	-	-
Krip. α	0.22	0.19	0.20	0.17	0.13

TABLE II: Top three most correlated action units (AU) and five scales. ρ : Spearman ranking correlation coefficient. Only correlation coefficient superior to 0.05 are included ($p < 0.0001$).

scale	AU	ρ	AU	ρ	AU	ρ
Interest	AU12	0.099	AU6	0.071	AU7	0.062
Curiosity	AU12	0.112	AU6	0.074	AU2	0.070
Coping	AU5	0.093	-	-	-	-
Novelty	AU23	0.119	AU5	0.105	AU14	0.088
Complexity	AU5	0.113	-	-	-	-

AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26 and AU45. OpenFace tracks 68 landmarks on the face. After rotating the faces to a frontal position, we registered them to a normalized face via rigid transformation found by Procrustes analysis on shapes from each frame. We calculated the Euclidean distance of 47 dynamic landmarks from the center as features. The following seven functionals were applied to the features in each trial for pooling: mean, standard deviation, median, maximum, minimum, first and third quartiles. This resulted in a feature vector with 329 elements for each trial.

We calculated the Pearson ranking correlation between the action units (averaged over each trial) and the ratings. The top three highly correlation action units with each scale are given in Table II. Interest has the highest correlation with action units activated with action units associated with smile (AU6 + AU12). Curiosity has a similar pattern to interest in addition to AU2 (brow raiser) which is also associate with novelty. Coping potential and complexity are correlated with AU5 which is the eye lid raiser. Novelty is surprisingly associated with AU14 (dimpler) and AU23 lip tightener in addition to AU5. It is also worth noting that, the model we use do not have a high performance in detecting all action units and these results are by no means comparable to the studies in psychology with manual coding.

V. EXPERIMENTAL RESULTS

We used an ensemble regression model, random forests, with 200 trees with minimum leaf size of five for detecting the level of interest, curiosity and appraisals. We used 20-folding cross-validation strategy for evaluating the regression results on five different scales, namely, interest, curiosity, coping potential, novelty and complexity. We also per-

TABLE III: Regression results on interest, curiosity, coping potential, complexity and novelty. For RMSE the maximum value is 1. For intra-participant (per participant) results, median and standard deviation values are given.

Scale	Spearman $\rho \uparrow$	CCC \uparrow	RMSE \downarrow
Inter-participant			
Interest	0.32	0.21	0.31
Curiosity	0.32	0.19	0.32
Coping potential	0.47	0.27	0.24
Novelty	0.36	0.24	0.33
Complexity	0.44	0.28	0.28
Intra-participant			
Interest	0.22(0.19)	0.11(0.16)	0.31(0.08)
Curiosity	0.17(0.16)	0.11(0.12)	0.31(0.08)
Coping potential	0.11(0.16)	0.06(0.10)	0.24(0.5)
Novelty	0.11(0.18)	0.06(0.12)	0.33(0.07)
Complexity	0.10(0.14)	0.06(0.09)	0.27(0.06)

formed the regression on each participants' responses using a leave-one-out cross validation. Results were evaluated using Spearman ranking correlation due to the ordinal nature of the scores). We also report root-mean-square error and concordance correlation coefficient (CCC). The regression evaluation results are given in Table III.

The inter-participant results are inferior compared to the intra-participant ones. This is partly due to the lower number of training data in that case. Overall, Coping potential and complexity were detected with higher accuracy compared to interest. This is partly due to the differences between the expressions associated with interest, e.g., smile and eyes open. It appears that the expressions of appraisals such as perceived complexity is more consistent between participants. However, within participant results for interest are superior, a sign of superior consistency for the expression of interest compared to appraisals within participants.

VI. CONCLUSIONS AND OUTLOOK

In this work, we conducted an experiment with the goal of assessing visual interest and its related components. Analysis on the action units showed that interest is related to eye opening and smile, which are signs of novelty and pleasantness. Indeed, in our preliminary study we found a correlation between pleasantness and interest for such a context. The promising results on the detection of interest and appraisals demonstrate that it is feasible to detect appraisals from facial expressions. In the future, features extracted by deep learning models shall be used for such a purpose.

VII. ACKNOWLEDGMENTS

We would like to thank Danny Dukes for discussions and Michael Gygli for his help with the videos.

REFERENCES

- [1] P. J. Silvia, "InterestThe Curious Emotion," *Current Directions in Psychological Science*, vol. 17, pp. 57–60, 2008.
- [2] P. J. Silvia and P. J., *Exploring the Psychology of Interest*. Oxford University Press, apr 2006.
- [3] P. J. Silvia, R. A. Henson, and J. L. Templin, "Are the sources of interest the same for everyone? using multilevel mixture models to explore individual differences in appraisal structures," *Cognition and Emotion*, vol. 23, no. 7, pp. 1389–1406, 2009.
- [4] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.
- [5] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05*. New York, New York, USA: ACM Press, Nov. 2005, p. 677.
- [6] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, and L. Van Gool, "The Interestingness of Images," in *The IEEE International Conference on Computer Vision (ICCV)*, 2013.
- [7] J. Reeve, "The face of interest," *Motivation and Emotion*, vol. 17, no. 4, pp. 353–375, Dec. 1993.
- [8] K. R. Scherer and H. Ellgring, "Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal?" *Emotion*, vol. 7, no. 1, pp. 113–30, 2007.
- [9] M. Mortillaro, B. Meuleman, and K. R. Scherer, "Advocating a Componential Appraisal Model to Guide Emotion Recognition," *International Journal of Synthetic Emotions*, vol. 3, no. 1, pp. 18–32, jan 2012.
- [10] A. Kapoor, R. Picard, and Y. Ivanov, "Probabilistic combination of multiple modalities to detect interest," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3. IEEE, 2004, pp. 969–972 Vol.3.
- [11] D. Gatica-Perez, I. McCowan, D. Zhang, and S. Bengio, "Detecting group interest-level in meetings," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE Press, 2005, pp. 489–492.
- [12] E. Kurdyukova, S. Hammer, and E. André, "Personalization of content on public displays driven by the recognition of group context," in *Ambient Intelligence*, ser. Lecture Notes in Computer Science, F. Paterno, B. Ruyter, P. Markopoulos, C. Santoro, E. Loenen, and K. Luyten, Eds. Springer Berlin Heidelberg, 2012, vol. 7683, pp. 272–287.
- [13] B. Schuller, R. Müller, F. Eyben, J. Gast, B. Hörnler, M. Wöllmer, G. Rigoll, A. Höthker, and H. Konosu, "Being bored? Recognising natural interest by extensive audiovisual integration for real-life application," *Image and Vision Computing*, vol. 27, no. 12, pp. 1760–1774, 2009.
- [14] M. Soleymani, "The quest for visual interest," in *Proceedings of the 23rd Annual ACM Conference on Multimedia*. ACM, 2015, pp. 919–922.
- [15] M. Gygli, Y. Song, and L. Cao, "Video2GIF: Automatic Generation of Animated GIFs from Video," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [16] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, mar 2016, pp. 1–10.
- [17] T. Baltrusaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic Action Unit detection," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, may 2015, pp. 1–6.